

# PROBLEMS WITH MEASURING AND CLASSIFYING MENTAL DISORDERS

ISSUES IN CLINICAL AND ABNORMAL  
PSYCHOLOGY NO.3

KEVIN BREWER

ISBN: 978-1-904542-54-4

This document is produced under two principles:

1. All work is sourced to the original authors. The images are all available in the public domain (most from [http://commons.wikimedia.org/wiki/Main\\_Page](http://commons.wikimedia.org/wiki/Main_Page)). You are free to use this document, but, please, quote the source (Kevin Brewer 2010) and do not claim it as you own work.

This work is licensed under the Creative Commons Attribution (by) 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/> or send a letter to Creative Commons, 171 2nd Street, Suite 300, San Francisco, California, 94105, USA.

2. Details of the author are included so that the level of expertise of the writer can be assessed. This compares to documents which are not named and it is not possible to tell if the writer has any knowledge about their subject.

Kevin Brewer BSocSc, MSc

An independent academic psychologist, based in England, who has written extensively on different areas of psychology with an emphasis on the critical stance towards traditional ideas.

Orsett Psychological Services,  
PO Box 179,  
Grays,  
Essex  
RM16 3EW  
UK

[orsettpsychologicalservices@phonecoop.coop](mailto:orsettpsychologicalservices@phonecoop.coop)

(<http://kmbpsychology.jottit.com>) or  
(<http://psychologywritings.synthasite.com/>)

## **CONTENTS**

	Page Number
1. How Many People are Mentally Ill?	4
2. Classification Systems for Mental Disorders	11
3. Ill, But Not Quite: Sub-Threshold Disorders	17
4. Measuring Hypothetical Constructs	22
5. Using Self-Report Questionnaires	24
6. Measuring Depression with Two Classic Interviewer-Scored Rating Scales	30
7. Are Childhood Mental Disorders on the Increase, Particularly Autism and ADHD?	35
8. Reliability and Validity in Measuring Mental Disorders	45
9. Eating Disorder Examination: An Investigator-Based Interview	49
10. New Mental Disorders: Pathological Grief Disorder	52
11. Adult Psychiatric Morbidity Survey 2007: Methodology	54

# **1. HOW MANY PEOPLE ARE MENTALLY ILL?**

- 1.1. Issues involved
- 1.2. Examples of studies
- 1.3. Example with schizophrenia
- 1.4. References

## **1.1. ISSUES INVOLVED**

There are many studies or reports giving figures for amount of mental illness or particular mental disorders, so the answer to the question in the title must be about finding the correct study. But the accurate measurement of mental disorders is difficult, and figures for the amount are not necessarily the true quantity. The purpose of this article is to highlight many of the problems involved in collecting data about mental illness.

Here are key variables when collecting such data.

1. The definition used - Official categories from the classification systems, like DSM-IV, or general language (eg: "unhappy", "blue" for depression).
2. The method of interviewing used in surveys - Face-to-face, postal, telephone, or via the Internet.
3. Who is reporting the information - Self-reported is most common, but also other-reported (eg: parents or teachers of children).
4. The sample used - It is not possible to question everybody in the population, so a smaller group (sample) is approached. How large is the sample? Is it representative of the population in terms of age, gender, social class, for example?
  - i) General population or community sample - Usually random selection of individuals from the general population.
  - ii) Specialist samples - Individuals in particular places; eg: GPs' patients, inpatients at hospitals, outpatients at clinics, prison inmates.
5. The use of official statistics - Governments collect vast amounts of data, but there are limitations to such figures, including willingness of individuals to give

information to official bodies, and definitions used.

6. The measurement of mental illness - The use of psychometric questionnaires or other questionnaires and survey instruments.

7. Behaviour measures - Rather than asking individuals, actual behaviour can be used as the measure; eg: consultation with GP, admission rates to hospital, prescription medication used.

8. Time frame used - Prevalence (total number of cases in population) can vary between lifetime (ever had), time limited (eg: last twelve months), point in time, or current. While incidence is the number of new cases in a certain period.

9. International comparisons - Many of the above issues make it difficult to compare figures between countries or areas.

10. Time comparisons - As the last point, but over time.

## **1.2. EXAMPLES OF STUDIES**

### **1. General population**

#### **a) Government-funded household survey**

The "Adult Psychiatric Morbidity, 2009" report (McManus et al 2009) collected data for England for the government (ie: classed as "official statistics").

The Clinical Interview Schedule-Revised (CIS-R) (Lewis et al 1992) was used for measuring mental disorders, and it has a cut-off score of 12 or more for diagnosis. Deverill and King (2009) reported that 15.1% of adults (aged 16 years or more) had a score of twelve or more. There was a gender difference - 18.4% of women had this score compared to 11.6% of men.

Comparisons across recent years in England, using CIS-R with 16-64 year-old adults, showed increased numbers of people scoring 12 or more: 16.4% (2007), 16.3% (2000), and 14.1% (1993) (Deverill and King 2009).

In terms of common mental disorders (CMDs) <sup>1</sup>, 16.2% was the prevalence figure for the last week in adults 16 years and older in 2007 (ie: when data collected). Women reported more than men again (19.7% vs 12.5%). Rates have also increased since 1993 (Deverill and King 2009). But the figures are affected by the fact that an individual can have more than one CMD.

#### b) Non-Western society

Nandi et al (2000) surveyed households in an area sixty kilometres from Calcutta, India using their own diagnostic categories in 1992. The rate per 1000 for depression was 74, and two for anxiety, for example.

Patel et al (2006) recruited 2166 women (18-50 years old) in the state of Goa, India for interview with the CIS-R. The rate for CMDs in the last twelve months was 1.8%.

The risk of CMDs was associated with poverty, being married, tobacco use, reporting chronic physical illness, and poor reproductive health (table 1.1).

RISK FACTOR	ODDS RATIO	ODDS RATIO = 1
Chronic physical illness	2.30	Good health
Tobacco use in past three months	3.23	No use
Married	6.02	Single
High income	0.41	Low income

(After Patel et al 2006)

Table 1.1 - Odds ratio of CMDs for certain variables.

#### c) Official diagnostic categories

Kessler et al (2005) reported the results from the National Co-morbidity Survey Replication (NCS-R) in the USA with a nationally representative sample and face-to-face interviews using DSM-IV criteria. Between February 2001 and April 2003, 9282 English-speaking adults (18 years and above) were interviewed.

The lifetime prevalence of disorders included 16.6% for major depressive disorder, 12.5% for specific phobia, and 28.8% for anxiety disorders. For any disorder the rate was 46.6%.

---

<sup>1</sup> CMDs include mixed anxiety and depressive disorder, generalised anxiety disorder, depressive episode, all phobias, obsessive-compulsive disorder, and panic disorder.

Kessler et al (2005) admitted that the figures may be underestimates because:

- The sampling frame excluded groups like homeless people who have a higher rate of mental illness than the general population;
- The reluctance of individuals with a history of mental illness to participate in such surveys;
- The under-reporting by respondents of embarrassing behaviours;
- The risk of recall failure for past behaviours.

Araya et al (2001a) found that 26.7% of 3870 adults in Santiago, Chile had any mental disorder using ICD-10 criteria, but women were twice as likely than men (35.2% vs 17.3%).

## 2. Specialist populations

### a) GP patients

Araya et al (2001b) compared the diagnosis of 815 consecutive patients in Santiago, Chile by GPs with the researchers' diagnosis using the CIS-R. There was 48% agreement in diagnosis with prevalence rates varying between 49% with CIS-R and 35% by GPs.

### b) Hospitals/clinics

For example, 7% of 800 children and adolescents at a Pennsylvania day-unit had adjustment disorders (Doan and Petti 1989) compared to 34% of US adolescent psychiatric inpatients (Greenberg et al 1995).

### c) Asylum seekers

Keller et al (2003) found clinically significant symptoms of anxiety in 77% of detained asylum seekers in the USA, while 86% had depression and 50% had Post-Traumatic Stress Disorder (PTSD).

### d) Virtual sample

Schlenger et al (2002) devised an Internet-based questionnaire on PTSD which was used with Americans two months after "9/11" (11 September 2001). Rates of PTSD

were 11.2% for New York city inhabitants and 4.3% in the whole of the USA.

### 1.3. EXAMPLE WITH SCHIZOPHRENIA

Schizophrenia is a greatly researched and discussed mental disorder such that, it is argued, that diagnosis and measurement is highly accurate. However, studies from around that world report different rates of this disorder.

Kirkbride et al (2006) calculated the rates of psychosis for adults (16-64 years old) using DSM-IV in three samples in England: south-east London (Lambeth and Southwark), central Bristol (both urban), and Nottinghamshire (urban, suburban, and rural) <sup>2</sup>. Table 1.2 shows the rates for schizophrenia and any psychosis.

AREA	ANY PSYCHOSIS	SCHIZOPHRENIA
Overall	32.1	11.7
London	49.4	20.1
Bristol	20.4	7.2
Nottinghamshire	23.9	7.6

Table 1.2 - Adjusted rates per 100 000 population.

While in Greater Sao Paulo, Brazil, Menezes et al (2007) found rates of 15.8 per 100 000 for psychosis, and 5.9 for "affective psychosis" <sup>3</sup> among adults 18-64 years using DSM-IV criteria.

Fekadu et al (2004) reported one case of schizophrenia out of 1691 participants (16 years plus) among the Zay population living on islands in Lake Zeway, Ethiopia (figure 1.1) using ICD-10 <sup>4</sup>. In another area of Ethiopia, Butajira (rural area 110km south of capital) (figure 1.1), using ICD-10 again, among adults 15-49 years old, schizophrenia was 4.7 per 1000 (Alem et al 2009).

The question is why are the rates different between countries, and what does that say about the universality of schizophrenia. Here are a number of possibilities:

---

<sup>2</sup> The data were collected as part of the Aetiology and Ethnicity in Schizophrenia and Other Psychoses (AESOP) study.

<sup>3</sup> This is a rate of 7.9 for schizophrenia (The Academy of Medical Sciences 2008).

<sup>4</sup> The rate for bipolar disorder was 1.83% though.





Butajira  
area

Lake  
Zeway

(My labels on map produced by Central Intelligence Agency in public domain)

Figure 1.1 - Two areas in Ethiopia.

- Schizophrenia is universal, but data collection problems and issues explain differences in prevalence rates. So if methodology is improved, the rates will show similarities.
- Schizophrenia is universal, but manifests itself differently in cultures around the world (sometimes viewed as a problem, sometimes not). So the differences in prevalence rates are accurate measures for that culture.

- Schizophrenia is an entirely cultural phenomenon, and differences between cultures reflect that fact. Thus the prevalence rates are accurate.

#### 1.4. REFERENCES

Alem, A et al (2009) Clinical course and outcome of schizophrenia in a predominantly treatment-naïve cohort in rural Ethiopia Schizophrenia Bulletin 35, 3, 646-654

Araya, R et al (2001a) Common mental disorder in Santiago, Chile British Journal of Psychiatry 178, 228-233

Araya, R et al (2001b) "Patient knows best" - detection of common mental disorder in Santiago, Chile: Cross-sectional study British Medical Journal 331, 79-81

Deverill, C & King, M (2009) Common mental disorders. In McManus et al op cit

Doan, R.J & Petti, T.A (1989) Clinical and demographic characteristics of child and adolescent partial hospitalised patients Journal of the American Academy of Child and Adolescent Psychiatry 28, 66-69

Fekadu, A et al (2004) Bipolar disorder among an isolated island community in Ethiopia Journal of Affective Disorders 80, 1-10

Greenberg, W.M et al (1995) Adjustment disorder as an admission diagnosis American Journal of Psychiatry 152, 459-461

Keller, A.S et al (2003) Mental health of detained asylum seekers Lancet 362, 1721-1723

Kessler, R.C et al (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Co-morbidity Survey Replication Archives of General Psychiatry 62, 593-602

Kirkbride, J.B et al (2006) Heterogeneity in incidence rates of schizophrenia and other psychotic syndromes Archives of General Psychiatry 63, 3, 250-258

Lewis, A.G et al (1992) Measuring psychiatric disorder in the community: A standardised assessment for use by lay interviewers Psychological Medicine 22, 465-486

McManus, S et al (2009) (eds) Adult Psychiatric Morbidity, 2009: Results of a Household Survey London: National Centre for Social Research

Menezes, P.R et al (2007) Incidence of first-contact psychosis in Sao Paulo, Brazil British Journal of Psychiatry 191, sup 51, s102-s106

Nandi, D.N et al (2000) Psychiatric morbidity of a rural Indian community British Journal of Psychiatry 176, 351-356

Patel, V et al (2006) Risk factors for common mental disorders in women British Journal of Psychiatry 189, 547-555

Schlenger, W.E et al (2002) Psychological reactions to terrorist attacks: Findings from the National Survey of Americans' Reactions to September 11 Journal of the American Medical Association 288, 581-588

The Academy of Medical Sciences (2008) Challenges and Priorities for Global Mental Health Research in Low- and Middle-Income Countries London: The Academy of Medical Sciences

## **2. CLASSIFICATION SYSTEMS FOR MENTAL DISORDERS**

- 2.1. Two classification systems
- 2.2. Interviewer differences
- 2.3. References

### **2.1. TWO CLASSIFICATION SYSTEMS**

There are two classification systems for mental disorders commonly used. One produced by the World Health Organisation, where mental disorders are described as part of the "International Classification of Diseases" (chapter V) (currently in the tenth edition: ICD-10; WHO 1992). The other is the "Diagnostic and Statistical Manual of Mental Disorders" (fourth edition, currently: DSM-IV; APA 1994 or DSM-IV-TR; APA 2000) from the American Psychiatric Association. Both systems make use of diagnostic categories (table 2.1).

ARGUMENTS FOR	ARGUMENTS AGAINST
<ul style="list-style-type: none"><li>1. Diagnostic categories can be standardised to aid use by different psychiatrists in different places at different times.</li><li>2. They guide research and treatment.</li><li>3. They "help to bring order to chaos" (Furnham 2001).</li><li>4. They are more objective than diagnosis based on personal opinion (ie: without them).</li><li>5. They help sufferers to understand what they suffering with.</li></ul>	<ul style="list-style-type: none"><li>1. Diagnostic categories "depersonalises and dehumanises the person to whom the label is attached" (Furnham 2001).</li><li>2. They appear objective when they are far from it, and do include personal opinion.</li><li>3. The process of labelling can stigmatise individuals as well as produce a self-fulfilling prophecy.</li><li>4. Individual behaviour may be forced to fit the diagnostic category as in the Procrustean bed (Furnham 2001).</li><li>5. They have become reified (ie: seen as "real" when only descriptive aids).</li></ul>

Table 2.1 - Main arguments for and against diagnostic categories.

It is assumed that the descriptions of mental disorders "are at least similar if not identical" in both (Peters et al 1999).

However, there are examples which show that the two classification systems are not completely

interchangeable. This is usually studied by comparing the same individuals diagnosed with a particular disorder or not with both classification systems.

Peters et al (1999) used Post-Traumatic Stress Disorder (PTSD) in such a study. 1364 community volunteers in Australia were interviewed face-to-face by trained lay interviewers (ie: non-psychiatrists) using the Composite International Diagnostic Interview (CIDI) (WHO 1997). The CIDI can be used with either ICD-10 or DSM-IV diagnostic criteria.

The majority of the interviewees (1264) were not diagnosed as suffering from PTSD in the last twelve months on both ICD-10 and DSM-IV. Of the remaining one hundred individuals, 35 were diagnosed with PTSD by both criteria. This left a disagreement over sixty-five individuals with diagnosis of PTSD more likely with ICD-10 than DSM-IV (table 2.2) (figure 2.1).

	DSM-IV: No PTSD	DSM-IV: PTSD
ICD-10 DCR *: No PTSD	<b>1264</b>	6
ICD-10 DCR: PTSD	59	<b>35</b>

(Bold = agreement; \* ICD-10 DCR - WHO 1993)

(After Peters et al 1999)

Table 2.2 - Number of individuals and diagnostic agreement for PTSD.

In another study in Australia, Slade and Andrews (2001) compared the diagnosis of Generalised Anxiety Disorder (GAD) using CIDI (box 2.1) among 10 641 people in the Australian National Survey of Mental Health and Well-Being (NSMHWB). The majority of individuals were negative on both classification systems. This left 475 individuals of which 123 were diagnosed with GAD in the last twelve months in both cases (table 2.3) (figure 2.2).

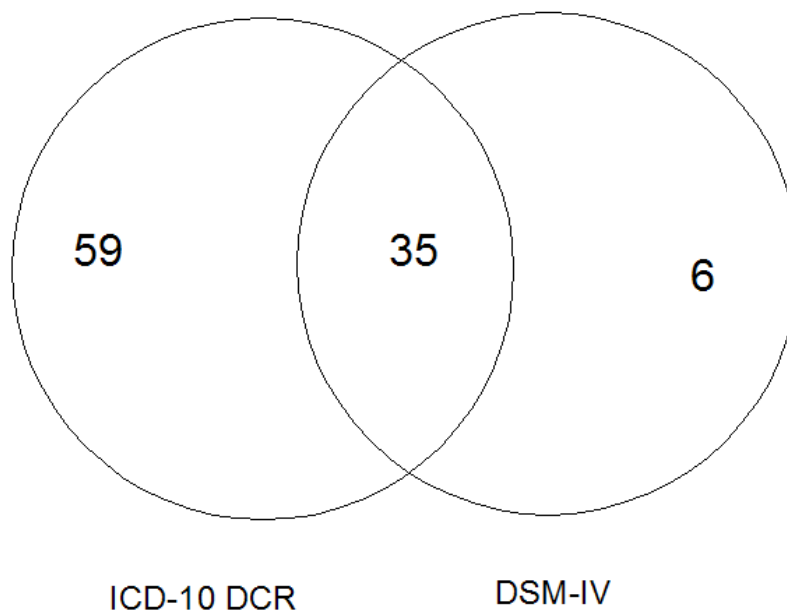


Figure 2.1 - Agreement and disagreement over diagnosis of PTSD.

	DSM-IV: No GAD	DSM-IV: GAD
ICD-10: No GAD	<b>10 166</b>	151
ICD-10: GAD	201	<b>123</b>

(Bold = agreement)

(After Slade and Andrews 2001)

Table 2.3 - Number of individuals and diagnostic agreement for GAD.

D63. The next questions are about longer periods of feeling worried, tense, or anxious. In the past 12 months, did you have a period of a month or more when most days you felt worried or tense or anxious about everyday problems such as work or family?

- YES
- NO
- DON'T KNOW
- REFUSE

D63.1. Did that period go on for at least six months?

- YES
- NO
- DON'T KNOW
- REFUSE

D63.2. How many months out of the last 12 did you feel worried or tense or anxious most days?

\_\_\_\_\_ MONTHS

D63.3 During (that/those) month(s), were you worried, tense, or anxious every day, nearly every day, most days, about half the days, or less than half the days?

1. EVERY DAY
2. NEARLY EVERY DAY
3. MOST DAYS
4. ABOUT HALF THE DAYS
5. LESS THAN HALF THE DAYS
6. DON'T KNOW
7. REFUSE

Box 2.1 - Examples of CIDI questions for GAD.

In both studies, the differences in diagnosis could be linked to specific criteria, but also to the fact that hypothetical (or unseen) constructs are being measured, and the language used. For example, the use of terms like "excessive" for worry and anxiety or symptoms causing "clinically significant distress".

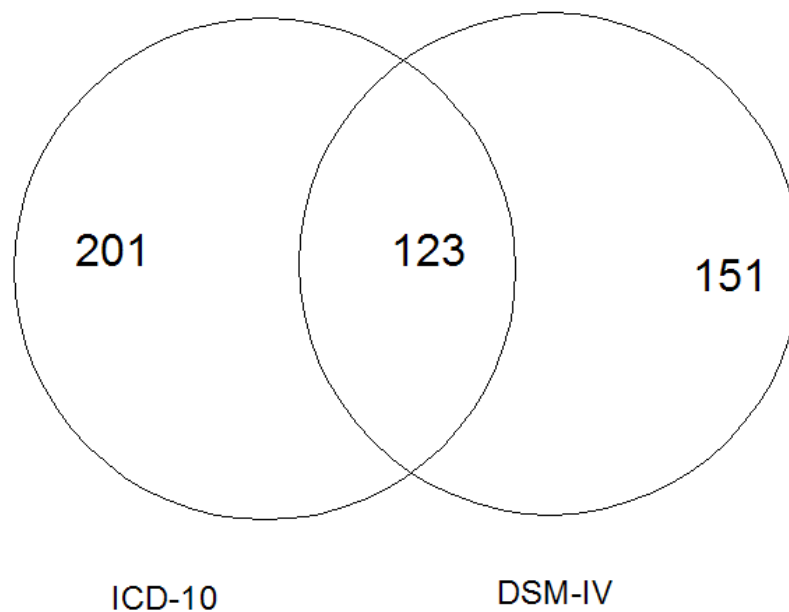


Figure 2.2 - Agreement and disagreement over diagnosis of GAD.

Whether the differences in diagnostic criteria influence actual diagnosis of mental disorders is open to debate because both systems emphasise the important features of the core mental disorders, but it will "undoubtedly hinder the comparison of prevalence rates from epidemiological surveys" (Slade and Andrews 2001).

## 2.2. INTERVIEWER DIFFERENCES

Even when standardised interviews for the same diagnostic system are used, there will be differences in diagnosis due to the interviewer, and, in particular, how they interpret and code responses. For example, the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders (SCID) (Jenca et al 1994) tends to use a "clinical impression" (ie: the interviewer's judgments) rather than the literal answers to questions.

Differences in diagnosis is a challenge to the reliability of the process. This can occur because:

- Differences between interviewers in the threshold for diagnosis or for severity;
- Different interviewers use the same threshold, but vary

in their accuracy of application of criteria for this threshold. This can produce "false positives" (non-cases diagnosed as cases) and "false negatives" (cases diagnosed as non-cases).

Grayson et al (1996) compared the diagnosis of post-traumatic stress disorder (PTSD) among 641 Australian Army Vietnam War Veterans by counsellors from the Vietnam Veterans Counselling Service, the research team, and psychologists in the Australian Army Psychology Corp (some were trained counsellors and some were not). The interviewers used a specially adaptive Australian version of SCID (AUSCID) and the Diagnostic Interview Schedule (DIS) (Robins et al 1981). The latter is more structured.

Using the DIS, lifetime prevalence of combat-related PTSD was found among 11.7% of the sample, but 20.7% with the AUSCID. Female counsellors diagnosed significantly higher rates of PTSD than female non-counsellors, and male counsellors and non-counsellors, and especially with the AUSCID. That difference could have been the veterans each group saw. But even when the characteristics of the veterans were controlled, the difference still stood.

## 2.3. REFERENCES

APA (2000) Diagnostic and Statistical Manual of Mental Disorders (4th ed - text revision) - DSM-IV-TR Washington DC: American Psychiatric Association

Furnham, A (2001) All in the Mind (2nd ed) London: Whurr

Grayson, D.A et al (1996) Interviewer effects on epidemiologic diagnoses of posttraumatic stress disorder American Journal of Epidemiology 144, 9, 589-597

Jenca, A et al (1994) New versions of World Health Organisation instruments for the assessment of mental disorders Acta Psychiatrica Scandinavica 90, 73-83

Peters, L et al (1999) A comparison of ICD-10 and DSM-IV criteria for Post-Traumatic Stress Disorder Journal of Traumatic Stress 12, 2, 335-343

Robins, L.N et al (1981) National Institutes of Mental Health. Diagnostic Interview Schedule: Its history, characteristics, and validity Archives of General Psychiatry 38, 381-389

Slade, T & Andrews, G (2001) DSM-IV and ICD-10 generalised anxiety disorder: Discrepant diagnoses and associated disability Social Psychiatry and Psychiatric Epidemiology 36, 45-51

WHO (1992) ICD-10 Classification of Mental and Behavioural Disorders - Clinical Descriptions and Diagnostic Guidelines Geneva: World Health Organisation

WHO (1993) ICD-10 Classification of Mental and Behavioural Disorders - Clinical Descriptions and Diagnostic Guidelines - Diagnostic Criteria for Research Geneva: World Health Organisation

WHO (1997) Composite International Diagnostic Interview, version 2.0 Geneva: World Health Organisation



### **3. ILL, BUT NOT QUITE: SUB-THRESHOLD DISORDERS**

The diagnosis of mental disorders is usually based upon a cut-off point on a continuum. Beyond that point the individual is classified as suffering from the mental disorder, and below that they are "normal" (not suffering from the mental disorder).

For example, with a scale of 0-10 for a particular behaviour, it is decided that seven and above constitutes evidence of the disorder. But there will be a difference between individuals scoring 5 or 6 and 1 or 2, though both are classed as "healthy". There is interest in the former group, known as "sub-threshold", "minor", or "sub-clinical" forms of the mental disorder (Batelaan et al 2006) (figure 3.1).

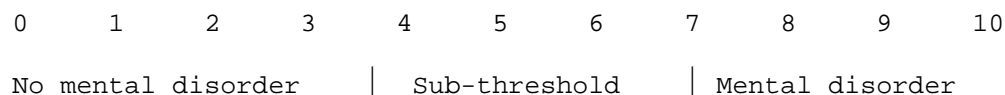


Figure 3.1 - Illustration of sub-threshold concept of mental disorders.

But how to deal with the concept of sub-threshold disorder? Lowering the threshold for full diagnosis of the disorder will increase the number of sufferers, and extends "abnormality" further into everyday life. Furthermore, a new lower sub-threshold is created. In a sense, this has happened already with the "pathologising of everyday life" and less tolerance of any signs of "unhappiness" (depression), for example.

An alternative is the double threshold concept (Helmchen and Linden 2000). One threshold defines mental disorder and the other mental health producing three groups - healthy, mildly ill, and ill (figure 3.2).

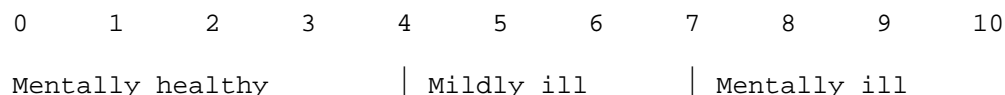


Figure 3.2 - Double threshold concept.

Batelaan et al (2006) applied the idea of a double threshold to panic disorder using data from the Netherlands Mental Health Survey and Incidence Study (NEMESIS) (Bijl et al 1998).

The Composite International Diagnostic Interview

(CIDI) for DSM-III-R (Dutch version; Smeets and Dingemans 1993) <sup>5</sup> was used to diagnose 7076 adults into three groups - no panic disorder (NPD) (n = 6770), panic disorder (PD) (n = 165), and sub-threshold panic disorder (sub-PD) (n = 141).

Panic disorder is diagnosed by four attacks within a month or one attack followed by at least a month of fear of another attack, and the panic attack has at least four of thirteen symptoms (eg: dizziness, sweating, choking). Diagnosis of sub-threshold panic disorder was at least one "sudden experience of intense fear in the year prior to the interview, in a situation in which most people would not be afraid" involving at least four of the thirteen symptoms (figure 3.3).

NPD	Sub-PD	DSM-III-R PD
No attacks in last year	One attack in last year	Four attacks within month

Figure 3.3 - Three groups of panic disorder sufferers.

On average, the sub-PD group experienced 7.5 of the thirteen symptoms of panic disorder compared to 9.4 for the PD group.

Analysis of the demographic variables found greater risk for panic disorder and sub-threshold panic disorder in a number of cases, like being female, while "not living with a partner", for example, was a stronger risk for sub-threshold than panic disorder (table 3.1).

---

<sup>5</sup> Example of questions on CIDI for panic disorder:

- Another kind of attack is when all of a sudden your heart begins to race, or you feel dizzy or faint, or you can't catch your breath. I'm not talking about a heart attack or some other attack caused by physical illness or medication or drugs, but about an attack that occurs for no apparent physical reason, just out of the blue. Have you ever had an attack like this?
- Have you had an attack like this in the past 12 months?
- In the past 12 months was there a month or more when you avoided certain situations or changed your everyday activities because of fear of the attacks?

VARIABLE	SUB-THRESHOLD	PANIC DISORDER
Female	1.87	3.37
Less than 44 years old	<b>2.09</b>	1.22
Urban living	1.40	1.67
Not living with partner	<b>2.05</b>	1.64
Not working	1.34	2.93
Low income	2.50	3.09
Low self-esteem	3.70	4.54

(After Batelaan et al 2006)

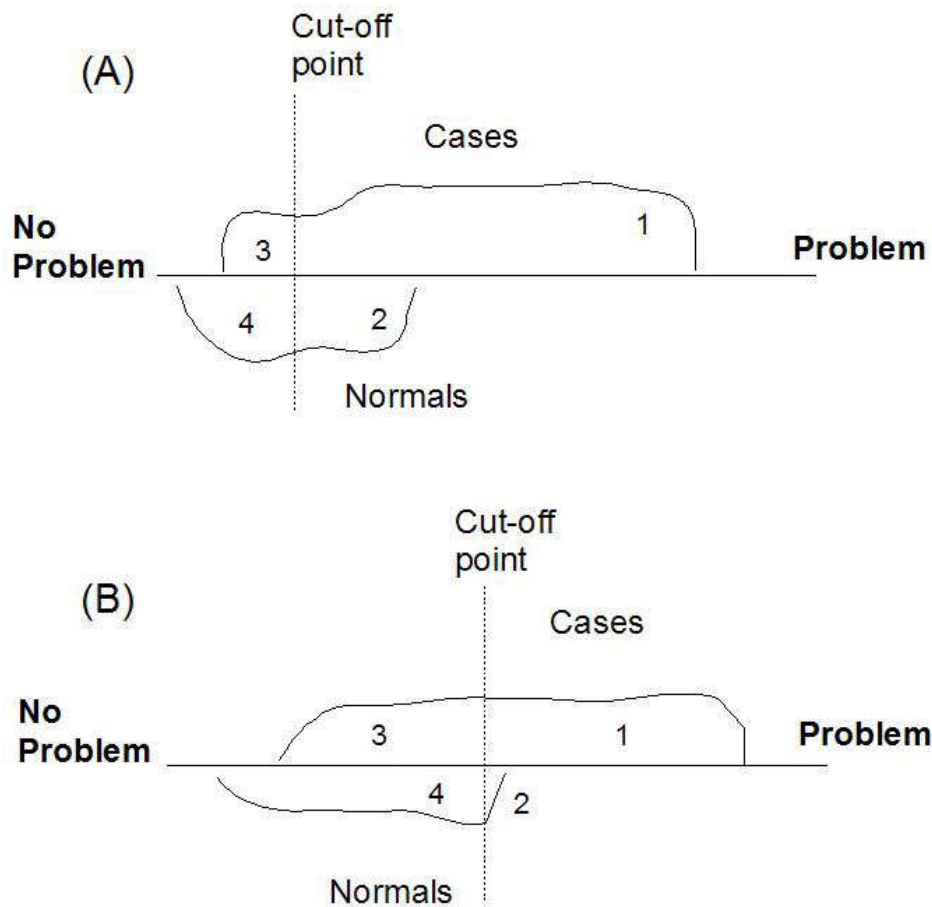
Table 3.1 - Odds ratio of sub-threshold and panic disorder compared to NPD group for selected variables.

The cut-off point can influence the accuracy of diagnosis of all cases. Figure 3.4 gives an example of a high and a low cut-off point.

In case (A) a low cut-off point covers most of the "true positives" <sup>6</sup> (actual cases) and misses few ("false negatives"), but it involves a lot of "false positives" (individuals diagnosed with a disorder when they do not have it). While case (B), a high cut-off point produces few "false positives", but misses many actual cases ("false negatives").

---

	<u>HAS DISORDER</u>	
	YES	NO
<u>DIAGNOSED WITH DISORDER</u>		
YES	true positive	false positive
NO	false negative	true negative



(A) = low cut-off point  
 (B) = high cut-off point  
 1 = true positives  
 2 = false positives  
 3 = false negative  
 4 = true negative

(Based on Fombonne 2002)

Figure 3.4 - Cut-off points and accuracy of diagnosis of all cases.

## REFERENCES

Batelaan, N et al (2006) Thresholds for health and thresholds for illness: Panic disorder versus sub-threshold panic disorder Psychological Medicine 37, 247-256

Bijl, R.V et al (1998) The Netherlands Mental Health Survey (NEMESIS):

Problems with Measuring and Classifying Mental Disorders. ISBN: 978-1-904542-54-4.  
 Issues in Clinical and Abnormal Psychology No.3. Kevin Brewer. 2010

Objectives and design Social Psychiatry and Psychiatric Epidemiology 33, 581-586

Helmchen, H & Linden, M (2000) Sub-threshold disorders in psychiatry: Clinical reality, methodological artifact, and the double-threshold problem Comprehensive Psychiatry 41, 1-7

Smeets, R.M.W & Dingemans, P.M.A.J (1993) Composite International Diagnostic Interview (CIDI), version 1.1 (in Dutch) Amsterdam/Geneva: World Health Organisation

## **4. MEASURING HYPOTHETICAL CONSTRUCTS**

"Measurement is a cornerstone of psychological research and practice. Measures of psychological constructs are used to test theories, to develop and evaluate applied intervention programs, and to assist practical psychologists in making treatment decisions" (Blanton and Jaccard 2006 p27).

Many constructs in psychology are hypothetical and cannot be directly observed. Instead, as with the example of depression, inferences are made from the individual's behaviour which is rated by another person using an inventory or test (ie: a scoring system). The "score is said to represent an individual's standing on a theoretical and unobservable psychological dimension" (Blanton and Jaccard 2006). This is quantified in some way - eg: amount, degree, or magnitude of behaviour.

The quantifying score will have a range of values (0-10, for example), and the term "metric" is used to refer to those numbers. These scores are representations of the behaviour not "true units of the unobserved psychological dimension" (Blanton and Jaccard 2006).

Furthermore, the score only makes sense in relation to other scores. "Until psychologists know what psychological reality surrounds the different scores on the scale, the response metric is arbitrary" (Blanton and Jaccard 2006). There is a need for external referents <sup>7</sup>.

Blanton and Jaccard (2006) felt that, in some cases, researchers draw meaning from the scores without external referents. They described two ways:

i) Meter reading - The score is assumed as a measure of the underlying dimension. For example, a high score on a depression inventory is seen as a high level of depression, and the opposite for a low score.

ii) Norming - Here raw scores on a questionnaire are converted into standardised scores (eg: z scores <sup>8</sup>) in relation to norms (like the mean) and meaning is inferred from the new score (box 4.1). A score on the depression inventory places the individual in the top 10% of scores, and thus is assumed to represent a high level of depression, for example.

---

<sup>7</sup> Measures and scoring systems can be reliable and valid, but still have "metric arbitrariness" (ie: lacking external referents or contexts to the scores) (Blanton and Jaccard 2006).

<sup>8</sup> Shows what proportion of population in normal distribution occurs at particular places. Z scores differ from standard deviation because relative to mean, while standard deviation shows unvarying percentage of population with a particular score in a normal distribution.

To reduce metric arbitrariness, the measurement scale should be grounded in meaningful events. Like on a ten-point scale, each number is linked to specific events (eg: a score of ten is associated with a suicide attempt) or in comparison to the general population (eg; 90% of individuals scored between 0-5). "It can be difficult and time consuming to conduct the research needed to make a metric less arbitrary" (Blanton and Jaccard 2006).

Data that does not have a normal distribution (ie: it is skewed) can be transformed into a normal distribution by logarithmic transformation. If the data is back transformed, the mean may be different to the original mean. This is now called the geometric mean (Bland and Altman 1996).

- Mean of data (skewed)
- Mean of log<sub>10</sub> transformed data (normal distribution)
- Mean of anti-log back transformed data (geometric mean)

Box 4.1 - Transforming data.

## REFERENCES

Bland, J.M & Altman, G (1996) Transformations, means, and confidence levels British Medical Journal 312, p1079

Blanton, H & Jaccard, J (2006) Arbitrary metrics in psychology American Psychologist 61, 1, 27-41

## **5. USING SELF-REPORT QUESTIONNAIRES**

- 5.1. Issues
- 5.2. Technical problems
- 5.3. Example of self-report questionnaire
- 5.4. References

### **5.1. ISSUES**

Questionnaires are a "fallible source of data" (Schwarz 1999) because the answers given can be influenced by aspects of the wording of the question. There are a number of issues to consider (Schwarz 1999).

#### **1. Making sense of the question**

Respondents attempt to make sense of the question asked both in terms of the literal meaning of the words and the pragmatic meaning (inferences about the questioner's intention) (Schwarz 1999).

In the latter case, contextual cues can influence the answers, like the title of the question or the name of the organisation carrying out the survey.

#### **2. Open versus closed questions**

Open-ended questions require the respondent to give their own answers, and it shows what they think about the question unprompted. But they may also spend time thinking about what the researcher wants and so edit out details.

Close-ended questions offer a limited choice of responses, and it is more focused for the researchers.

#### **3. Frequency versus time period**

Questions will usually refer to a period of time either directly (eg: "have you felt depressed in the last month"?) or indirectly through the frequency of an event (eg: "how frequently do you feel depressed"?). The responses offered are also important (eg: "less than once a month", "twice a week").

For example, Schwarz and Scheuring (1992; quoted in Schwarz 1999) compared the frequency of physical symptoms reported by psychosomatic patients offered "several times a day" to "twice a month or less" or "never" to "more than twice a month". In the first case, 62% chose "twice a month or less" and 39% the equivalent category in the second case.



While Schwarz et al (1985) found that over 60% of German respondents admitted to watching up to 2.5 hours of television a day when that was the lowest frequency (ie: six options from "up to 2.5 hours" to "more than 4.5 hours") compared to over 80% with low-frequency alternatives (ie: six options from "up to 0.5 hour" to "more than 2.5 hours").

"Essentially, respondents assume that the researcher constructs a meaningful scale, based on his or her knowledge of, or expectations about, the distribution of the behaviour in the 'real world'" (Schwarz 1999 pp97-98). Individuals will also make sense of their behaviour and answer subsequent questions in light of that.

#### 4. Size of the rating scale

Many questions offer a numerical scale for answers ranging in size (eg: 0-5, 0-10) or numbers (eg: -3 to +3).

Schwarz et al (1991) asked the question, "How successful would you say you have been in life?", and offered a ten-point scale in two different formats. In one format, -5 (not at all successful) to +5 (extremely successful), 34% of respondents chose zero or a minus score. While only 13% chose 0-5 on a scale of 0 (not at all successful) to 10 (extremely successful). On a scale with minus numbers the respondents felt able to choose in relation to negative characteristics (not successful), but not lower numbers if a positive scale (ie: 0-10).

#### 5. Question context

The preceding question can influence the current question. For example, Strack et al (1991) used the vague term "educational contribution" in a question to German students either following one about paying fees or about student grants. Very different responses were given depending upon the preceding question.

#### 6. General versus specific question

Lorenz and Ryan (1996) compared the effect on answers of a general question preceding or following related specific questions. The topic studied was satisfaction with community services in small US towns.

When the general question preceded the specific questions (GS), there were more positive responses than the other way around (SG) (table 5.1).

	GS	SG
Postal questionnaire	62	52
Telephone interviews	70	65

(After Lorenz and Ryan 1996)

Table 5.1 - Percentage of respondents choosing "good" and "very good" for general question about community services in survey <sup>9</sup>.

But when the question was about local government services, the effect was the opposite way around (table 5.2).

	GS	SG
Postal questionnaire	47	56
Telephone interviews	54	76

(After Lorenz and Ryan 1996)

Table 5.2 - Percentage of respondents choosing "good" and "very good" for general question about local government services.

Lorenz and Ryan (1996) interpreted the results based on the idea that with the general question first individuals respond with their "gut feeling", whereas the specific questions focus their attention. In the case of community services, the community is evaluated positively in the general question and the specific ones remind respondents of the negative issues (GS). In the case of government services, the initial evaluation is negative (ie: only recalling problems) while the specific questions can remind individuals of the positive aspects. However, the researchers admitted that this "post-hoc explanation seems convincing to us, but it isn't consistent with previous communication studies.." (p613).

This research also showed that respondents were more positive for the same questions in telephone interviews than in self-administered postal questionnaires.

Overall, it shows that responses to questions can be influenced by the order and nature of questions, as well as the method of asking used.

---

<sup>9</sup> "Please rate the overall quality of services and facilities located in (name of town)". Responses offered: very good, good, failing, poor. Specific questions related to nine services; eg: medical services, public schools.

## 5.2. TECHNICAL PROBLEMS

### 1. Missing data

When an individual fills in a questionnaire, they sometimes miss out items, either individual items/questions or whole sections. When there are omissions, it would make sense to discard the whole individual's responses, but this can reduce the sample size. Furthermore, this is wasteful if only one item is missed on a large survey and the rest are completed fully.

There are statistical techniques to deal with the missing data, like replacing the omission with the mean score for that item or variable (Fox-Wasylyshyn and El-Masri 2005). This can be more difficult if one item or variable is missed by many respondents. For example, Hartel (1976) recommended removing a variable from analysis if 15% of cases are missing, while Raymond and Roberts (1987) suggested inclusion until 40% of cases were missing.

### 2. Significance issues

The threshold for significance is conventionally set at  $p = 0.05$ , and this level is attributed to Fisher (1950). However, he did not assert the absolute threshold, rather the decision was up to the researcher: "If  $p$  is between 0.1 and 0.9 there is certainly no reason to suspect the hypothesis tested. If it is below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at 0.05.." (Fisher 1950 p80; quoted in Sterne and Davey Smith 2001 p226).

Significant results are more likely to be published than non-significant ones, but this means that "a host of purely chance findings will be published, as by conventional reasoning examining twenty associations will produce one result that is 'significant at  $p = 0.05$ ' by chance alone" (Sterne and Davey Smith 2001 p226).

### 3. Data carving

For purposes of analysis, continuous data is divided into categories. For example, the scores are divided into two groups either side of the median ("high", above the mean, and "low" below it). This gives equal numbers of scores in each group, but where do the median scores go? (Owen and Froman 2005).

How the division of data is made in this process of

"data carving" can influence the results.

### 5.3. EXAMPLE OF SELF-REPORT QUESTIONNAIRE

The Beck Depression Inventory (BDI) (Beck et al 1961) is a commonly used self-reported measure of depression (box 5.1).

**A (Mood)**

0. I do not feel sad; 1. I feel blue or sad; 2a. I am blue or sad all the time and I can't snap out of it; 2b. I am so sad or unhappy that it is very painful; 3. I am so sad or unhappy that I can't stand it

**B (Pessimism)**

0. I am not particularly pessimistic or discouraged about the future; 1a. I feel discouraged about the future; 2a. I feel I have nothing to look forward to; 2b. I feel that I won't ever get over my troubles; 3. I feel that the future is hopeless and that things cannot improve

**C (Sense of Failure)**

0. I do not feel like a failure; 1. I feel I have failed more than the average person; 2a. I feel I have accomplished very little that is worthwhile or that means anything; 2b. As I look back on my life all I can see is a lot of failures; 3. I feel I am a complete failure as a person (parent, husband, wife)

**D (Lack of Satisfaction)**

0. I am not particularly dissatisfied; 1a. I feel bored most of the time; 1b. I don't enjoy things the way I used to; 2. I don't get satisfaction out of anything any more; 3. I am dissatisfied with everything

**E (Guilty Feeling)**

**F (Sense of Punishment)**

**G (Self Hate)**

**H (Self Accusations)**

**I (Self-punitive Wishes)**

**J (Crying Spells)**

**K (Irritability)**

**L (Social Withdrawal)**

**M (Indecisiveness)**

**N (Body Image)**

**O (Work Inhibition)**

**P (Sleep Disturbance)**

**Q (Fatigability)**

**R (Loss of Appetite)**

**S (Weight Loss)**

**T (Somatic Preoccupation)**

**U (Loss of Libido)**

(Source: Beck et al 1961 Appendix)

Box 5.1 - Examples of statements from BDI.

## 5.4. REFERENCES

- Beck, A.T et al (1961) An inventory for measuring depression Archives of General Psychiatry 4, 53-63
- Fisher, R.A (1950) Statistical Methods for Research Workers London: Olive & Boyd
- Fox-Waslyshyn, S.M & El-Masri, M.M (2005) handling missing data in self-report measures Research in Nursing and Health 28, 488-495
- Hartel, B (1976) Minimising error variance introduced by missing data routines in survey analysis Sociological Methods and Research 4, 459-474
- Lorenz, F.O & Ryan, V.D (1996) Experiments in general/specific questions: Comparing results of mail and telephone surveys Proceedings of the Survey Research Methods Section, American Statistical Association 1996, 611-613
- Owen, S.V & Froman, R.D (2005) Why carve up your continuous data? Research in Nursing and Health 28, 496-503
- Raymond, M.R & Roberts, D.M (1987) A comparison of methods for treating incomplete data in selection research Educational and Psychological Measurement 47, 13-26
- Schwarz, N (1999) Self-reports: How the questions shape the answers American Psychologist 54, 2, 93-105
- Schwarz, N et al (1985) Response categories: Effects on behavioural reports and comparative judgments Public Opinion Quarterly 49, 388-395
- Schwarz, N et al (1991) Rating scales: Numeric values may change the meaning of scale labels Public Opinion Quarterly 55, 570-582
- Sterne, J.A.C & Davey Smith, G (2001) Sifting the evidence - what's wrong with significance tests? British Medical Journal 322, 226-231
- Strack, F et al (1991) Semantic and pragmatic aspects of context effects in social and psychological research Social Cognition 9, 111-125

## **6. MEASURING DEPRESSION WITH TWO CLASSIC INTERVIEWER-SCORED RATING SCALES**

- 6.1. Introduction
- 6.2. Hamilton Depression Rating Scale
  - 6.2.1. Evaluation
- 6.3. Montgomery-Asberg Depression Rating Scale
  - 6.3.1. Evaluation
- 6.4. General problems with rating scales
- 6.5. References

### **6.1. INTRODUCTION**

Depression is measured by a number of rating scales of which the Hamilton Depression Rating Scale and the Montgomery-Asberg Depression Rating Scale are two classics.

### **6.2. HAMILTON DEPRESSION RATING SCALE (HDRS)**

The HDRS (Hamilton 1960, 1967) has 17 items used to rate the severity of depression in the last week (box 6.1). It is used by an interviewer with the Structured Interview Guide for HDRS (SIGHD). There is a self-report version, the Carroll Rating Scale for Depression, which uses yes/no statements (Picardi 2009).

<ul style="list-style-type: none"><li>• Guilt</li></ul> <p>0 Absent 1 Mild or trivial 2) 3) Moderate 4 Severe</p> <ul style="list-style-type: none"><li>• Suicide</li></ul> <p>0 Absent 1 Mild or trivial 2) 3) Moderate 4 Severe</p>	<ul style="list-style-type: none"><li>• Agitation</li></ul> <p>0 Absent 1 Slight or doubtful 2 Clearly present</p> <ul style="list-style-type: none"><li>• Depressed mood</li></ul> <p>0 Absent 1 Mild or trivial 2) 3) Moderate 4 Severe</p>
---	---

(Source: Hamilton 1960 Appendix I p61)

Box 6.1 - Example of items from HDRS.

### **6.2.1. Evaluation**

1. Low reliability of some items: intra-class reliability varies from 0.46 - 0.99, and test-retest reliability of items varies from 0.00 - 0.85. But overall test-retest reliability varies from 0.81 - 0.98, and inter-rater reliability 0.82 - 0.98 <sup>10</sup>. Internal reliability adequate: 10 studies produced a range of correlations, 0.46 - 0.97 (Bagby et al 2004).
2. Content validity poor, but convergent and divergent validity adequate in studies (Bagby et al 2004).
3. It is criticised for multidimensionality (ie: measuring more than one dimension of behaviour) (Picardi 2009).
4. It gives excessive weight to somatic and anxiety features of depression (Picardi 2009).
5. It is unable to distinguish between patients with different depressive symptom profiles (Demyttenaere and De Fruyt 2003).
6. Bipolar depression does not overlap with unipolar depression, so HDRS of limited use with the former (Picardi 2009).
7. Designed for use with individuals already diagnosed with depression (Hamilton 1960).

### **6.3. MONTGOMERY-ASBERG DEPRESSION RATING SCALE (MADRS)**

The MADRS (Montgomery and Asberg 1979) has ten items for interviewer use (box 6.2) <sup>11</sup>. It assesses the last week, and is sensitive to change during treatment. A self-report version (MADRS-S) has nine items because "apparent sadness" cannot be self assessed (Picardi 2009).

---

<sup>10</sup> Two independent raters scoring the same patient simultaneously is ideal (Hamilton 1960).

<sup>11</sup> It was derived from the 65-item Comprehensive Psychopathology Rating Scale (Picardi 2009).

ITEMS:	Reduced appetite
1. Apparent sadness	Representing the feeling of a loss of appetite compared with when well. Rate of loss of desire for food or the need to force oneself to eat.
2. Reported sadness	
3. Inner tension	
4. Reduced sleep	
5. Reduced appetite	
6. Concentration difficulties	
7. Lassitude	0 Normal or increased appetite
8. Inability to feel	1
9. Pessimistic thoughts	2 Slightly reduced appetite
10. Suicidal thoughts	3
	4 No appetite. Food is tasteless
	5
Suicidal thoughts	6 Needs persuasion to eat at all
0 Enjoys life or takes it as it comes	
1	
2 Weary of life. Only fleeting suicidal thoughts	Reported sadness
3	
4 Probably better off dead. Suicidal thoughts are common, and suicide is considered as a possible solution, but without specific plans or intention	0 Occasional sadness in keeping with the circumstances
5	1
6 Explicit plans for suicide when there is an opportunity. Active preparations for suicide	2 Sad or low but brightens up without difficulty
	3
	4 Pervasive feelings of sadness or gloominess. The mood is still influenced by external circumstances
	5
	6 Continuous or unvarying sadness, misery or despondency

(Source: Montgomery and Asberg 1979 Appendix pp387-389)

Box 6.2 - Example of items from MADRS.

### 6.3.1. Evaluation

1. It places greater emphasis on psychic (psychological) than somatic aspects of depression than HDRS, so less sensitive to drug side effects (Demyttenaere and De Fruyt 2003).

2. It lacks unidimensionality (Picardi 2009).

3. There is some dispute about the best cut-off point for remission of bipolar depression. Traditionally, this is a score of twelve or less, but Berk et al (2008) argued that five or less is better <sup>12</sup>.

4. Limited use with bipolar depression as with HDRS

<sup>12</sup> The HDRS uses a score of seven or less with bipolar depression (Berk et al 2008).



(Picardi 2009).

5. It is sensitive to change during treatment.

6. With fewer items than the HDRS, it is short and easy to apply in a clinical setting (Montgomery and Asberg 1979).

#### **6.4. GENERAL PROBLEMS WITH RATING SCALES**

Hamilton (1960) addressed the common problems with depression rating scales available at the time, but these problems are relevant today.

i) Depression rating scales are not useful with the general population because such individuals do not suffer from many of the symptoms of depression.

ii) Self-reported scales are of limited use with semi-literate, and with seriously ill patients.

iii) Many scales are devised for a specific context like in a hospital ward, and so are only useful in that context.

iv) Symptoms are not of equal importance in different mental disorders. For example, individuals with schizophrenia can show anxiety, but that symptom is less important compared to other symptoms, and compared to individuals with anxiety disorders.

Montgomery and Asberg (1979) commented upon the problems of long scales: "the presence of a large number of items that were scored in only a few patients would tend to introduce and increase the random error. More important, the ratings would be cumbersome and time-consuming to undertake. Unskilled raters might have difficulties in covering a large number of items in a single interview. Repeated asking of questions which appear irrelevant to the patient might also be detrimental to clinical rapport and reduce the validity of the information provided" (p385).

#### **6.4. REFERENCES**

Bagby, R.M et al (2004) The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? American Journal of Psychiatry 161, 2163-2177

Berk, M et al (2008) The empirical redefinition of the psychometric criteria for remission in bipolar disorder Journal of Affective Disorders 106, 153-158

Demyttenaere, K & De Fruyt, J (2003) Getting what you ask for: On the selectivity of depression rating scales Psychotherapy and Psychosomatics 72, 61-70

Hamilton, M (1960) A rating scale for depression Journal of Neurology, Neurosurgery and Psychiatry 23, 56-62

Hamilton, M (1967) Development of a rating scale for primary depressive illness British Journal of Social and Clinical Psychology 6, 278-296

Montgomery, S & Asberg, M (1979) A new depression scale designed to be sensitive to change British Journal of Psychiatry 134, 382-389

Picardi, A (2009) Rating scales in bipolar disorder Current Opinion in Psychiatry 22, 42-49

## **7. ARE CHILDHOOD MENTAL DISORDERS ON THE INCREASE, PARTICULARLY AUTISM AND ADHD?**

- 7.1. Autism
- 7.2. Attention deficit hyperactivity disorder
- 7.3. Reasons for the increased number of cases of childhood mental disorders
- 7.4. Making sense of the figures
- 7.5. References

### **7.1. AUTISM**

There is an increase in the incidence of autism being reported (Chakrabarti and Fombonne 2005). In the USA, official statistics showed an increase of over 600% between 1993 and 2003 (Lilienfeld and Arkowitz 2007).

The rates for autism spectrum disorders (ASD) in the USA were 0.1 - 0.4 per 1000 children in the 1980s and between 2.0 - 7.0 per 1000 children in the 1990s (Centers for Disease Control and Prevention 2007a). In 2000, the Autism and Developmental Disabilities Monitoring (ADDM) Network collected extensive data from six US states on 1252 children aged eight years old. The prevalence of ASDs ranged from 4.5 in West Virginia to 9.9 in New Jersey per 1000 children (mean = 6.7) (Centers for Disease Control and Prevention 2007a).

This was extended to fourteen sites and 407 578 children (Centers for Disease Control and Prevention 2007b). The mean rate was 6.6 per 1000 children (range 3.3 - 10.6).

This has led to the use of the term "epidemic". Fombonne (2003) highlighted concerns with the idea of an "autism epidemic":

- Recent studies use a wider concept of ASD which includes autism disorder, Asperger disorder, and pervasive developmental disorders, whereas older studies (eg: 1960s and 1970s) used a narrow definition of autism only. This latter definition does not include autism occurring with learning disabilities. Thus comparisons across time are difficult.
- Studies that find increases over time find different sized increases, and show the number of confounding variables involved in such studies.
- The use of measures like referrals to professionals can also change over time (ie: who is referred and why).

Grinker (2007) argued that the increase is "merely a

shift in the cultural conditions that change the way medical scientists do their work and how we perceive mental health" (Madsen 2007).

Atladottir et al (2007) used the data from 669 995 children born between 1st January 1990 and end of 1999 in Denmark (divided into five birth cohorts; eg: 1990-1), and a clinical diagnosis (ICD-10) of hyperkinetic disorder <sup>13</sup>, obsessive-compulsive disorder (OCD), Tourette syndrome, autism spectrum disorder and childhood autism between January 1995 and December 2004 (as recorded on the Danish National Psychiatry Register). The incidence of the disorders was calculated per 10 000 children.

There was a significant increase for each disorder except OCD:

- Hyperkinetic disorder - significant increase for each of four birth cohorts in the study;
- Autism spectrum disorder - significantly greater in the 1998-9 birth cohort compared to the 1994-5 one;
- Childhood autism - significantly higher in the 1998-9 cohort compared to 1994-5 and 1996-7 cohorts;
- Tourette syndrome - incidence significantly highest between 1990-1 and 1992-3 cohorts (Madsen 2007).

This study showed that it is not just the incidence of autism that is increasing in recent years but also others childhood disorders (Madsen 2007).

## **7.2. ATTENTION DEFICIT HYPERACTIVITY DISORDER (ADHD)**

Since the beginning of the twentieth century, "upwards of twenty different diagnostic labels have been used to categorise children who exhibit ..problematic behaviours" (Mayes and Rafalovich 2007 p436). Furthermore, "What is striking about the numerous terms used to describe these children is the fact that they have been used for essentially the same behaviour symptoms as those first outlined in 1902. And what is perhaps equally striking is that while these children have remained similar in terms of their description decade after decade, albeit under different diagnostic labels, the explanations offered for their condition have varied dramatically" (Mayes and Rafalovich 2007 p436).

In 1902, Sir George Frederick Still from King's College Hospital, London described in the "Lancet" 20 "behaviourally disturbed" children who "exhibited violent

---

<sup>13</sup> ICD-10 uses hyperkinetic disorder while DSM-IV prefers attention deficit hyperactivity disorder.

outbursts, wanton mischievousness, destructiveness and a lack of responsiveness to punishment" with a "quite abnormal incapacity for sustained attention, causing school failure even in the absence of intellectual retardation" (quoted in Mayes and Rafalovich 2007). The label given for these children was a "defect of moral control" due to "the manifestation of some morbid physical condition" (Still 1902 p1165).

In 1922, Alfred Tredgold offered the explanation for such children as mild brain damage, usually at birth, leading to "feeble-mindedness". After the influenza epidemic of 1918, "post-encephalitic behaviour disorder" was a term used to describe children who had survived the infection but showed problem behaviours (Mayes and Rafalovich 2007).

Subsequent terms for children's problem behaviour included "hyperkinetic impulse disorder" (Laufer et al 1957) where "hyperactivity is the most striking item", "minimal brain dysfunction" (MBD) (Clements and Peters 1962), and "hyperkinetic reaction of childhood" in DSM-II (APA 1968); eventually, attention deficit disorder (ADD) in DSM-III (APA 1980), and ADHD was added as a sub-type of ADD in DSM-III-R (APA 1987).

Seeing children's problem behaviour as a medical condition has allowed the use of drugs as a treatment beginning with Benzedrine (amphetamine) by Charles Bradley in the 1930s in the USA. Chlorpromazine (neuroleptic) was also tried in the mid-1950s. The now commonly used methylphenidate, "Ritalin" (brand name), was synthesised by Swiss pharmaceutical firm, J.R. Geigy in 1955 and licensed for use in the USA in 1961 (Mayes and Rafalovich 2007).

Schrag and Divoky (1975) were one of the first in the USA to argue that children were "being labelled with a dubious diagnosis", and being given a "chemical straitjacket" "to control the natural exuberance and activity of children who came into conflict with teachers or other school personnel" (Mayes and Rafalovich 2007 p451).

Conrad (1975) saw the label as a "means of social control": "By focusing on the symptoms and defining them as hyperkinesis we ignore the possibility that behaviour is not an illness but an adaptation to a social situation. It diverts our attention ..from seriously entertaining the idea that the 'problem' could be in the structure of the social system" (p19).

### 7.3. REASONS FOR THE INCREASED NUMBER OF CASES OF CHILDHOOD MENTAL DISORDERS

There are a number of possible reasons for the increase in cases of childhood mental disorders reported (table 7.1).

EXPLANATION FOR INCREASE	GENUINE INCREASE IN NUMBERS?
1. More children suffer today that in past	Yes; something distinct today is causing increase
2. Diagnosis process better today	No; many cases missed in past
3. Greater awareness today	No; many cases missed in past
4. Increasing number of categories of mental disorder	No; more behaviour is now classed as abnormal than in past
5. Changes in attitudes in society	No; society is less tolerant of behaviour classed as normal in past
6. Official policies encourage diagnosis	No; more children diagnosed to gain educational assistance, for example
7. Facet of data collection	No; methodological problems explain different number of cases

Table 7.1 - Possible reasons for increased number of cases of childhood mental disorders today.

REASON 1. There is a genuine increase in the numbers of children with the disorders. The figures accurately show that more children suffer today than in the past.

If the increase is genuine, then the search is on for aspects of society today that were not present in the past. This has produced a number of ideas related to the environment, like child vaccination, which have limited support (eg: Kaye et al 2001).

From the viewpoint of genetics, Comings (1996) proposed that modern society was selecting for genes linked to problem behaviours. Women who are more educated (in terms of years studied) have children later which is a risk factor for behaviour problems. Meanwhile, poorly educated individuals are having more children. This is not a widely accepted idea.

REASON 2. The increase in children with the disorders represents improvements in diagnosis and measurement. The increase is not genuine, but it is more cases being diagnosed today which were missed in the past. More children suffered in the past, but were not diagnosed with the conditions.

However, Shattuck (2006) noted evidence of "diagnostic substitution" that while the numbers of children with autism increased, the diagnoses of learning disabilities decreased in the same period in the USA (1994-2003). In other words, there was no overall increase in children with problems, the distribution of diagnoses had changed, and the focus was upon the increased categories.

Shattuck used the US Department of Education's "Special Education Child Counts" of children aged 6-11 years. The recorded rates of autism increased from 0.6 per 1000 children in 1994 to 3.1 in 2003, but the rates of learning disabilities declined by 8.3 per 1000 children for the same period (47.9 to 39.6 per 1000).

Also some criteria for diagnosis are subjective; eg: "often has difficulty organising tasks or activities" (ADHD; DSM-IV; Schneider and Eisenberg 2006).

REASON 3. Linked to the last point, today there is a greater awareness of the different childhood problems by parents, educators, and clinicians. This is sometimes called the "Rain Man Effect" after the film (Lilienfeld and Arkowitz 2007).

So behaviours that would have been missed or seen as normal are now spotted as symptoms of a disorder. Again there is not a genuine increase in numbers, merely a greater recognition now.

Pharmaceutical companies and their marketing of drugs through direct-to-consumer advertising (in the USA) has aided this process. Such companies benefit from increased diagnostic rates as most cases will be prescribed drugs, and thus sales will increase.

ADHD is more often first suggested by school teachers or other school personnel (52.4% of cases) or parents (30.2%) than medical professionals (14.4%) (Sax and Kautz 2003), which makes it open to a number of influences. Schneider and Eisenberg (2006) analysed data from the Early Childhood Longitudinal Survey - Kindergarten Cohort (ECLS-K) which is a nationally representative sample of five year olds in the USA. The estimated prevalence of ADHD was 5.44% based on parents' reports.

Schneider and Eisenberg (2006) were interested in differences in prevalence of ADHD depending on child, family, and school characteristics. Table 7.2 lists the most common variables. This study showed how different variables can influence the diagnosis of ADHD, particular in terms of increased diagnosis.

CHILD CHARACTERISTICS	FAMILY CHARACTERISTICS	SCHOOL CHARACTERISTICS
<ul style="list-style-type: none"> <li>• Male</li> <li>• White</li> <li>• Summer birth</li> </ul>	<ul style="list-style-type: none"> <li>• Mother older than 18 and younger than 38 years at birth</li> <li>• Not living with biological parents</li> <li>• Father and mother lower education</li> <li>• Lowest income quintile</li> </ul>	<ul style="list-style-type: none"> <li>• Older teacher</li> <li>• Non-White teacher</li> <li>• School under strict performance regime</li> </ul>

Table 7.2 - Characteristics associated with diagnosis of ADHD.

REASON 4. The number of children diagnosed with a mental illness has increased because the number of categories of mental disorders has increased in recent years.

For example, DSM-III (APA 1980) contains two categories relevant to autism, while DSM-IV (APA 1994) has five (Lilienfeld and Arkowitz 2007).

Overall, DSM-II contained 180 discrete disorders, DSM-III 265, DSM-III-R 292, and DSM-IV has 297 categories (Shorter 1997) or 330 if the appendices are included (Stone 1998).

The increase in the number of categories of mental disorders goes hand in hand with the "pathologising" of everyday behaviour, the increasing power of psychiatry and of biological psychiatry (Kutchins and Kirk 1997).

Also the criteria for diagnosis within each category has expanded (or loosened, depending upon how you look at it). Diagnosis of autism in 1980 (DSM-III) required evidence of all six criteria, but in 1994 (DSM-IV) it was eight of sixteen criteria (Lilienfeld and Arkowitz 2007).

Among the criteria, "a pervasive lack of responsiveness to other people" was required in 1980, but only "a lack of spontaneous seeking to share... achievements with other people" in 1994. While "bizarre responses to various aspects of the environment" (1980) became "persistent preoccupation with parts of objects" (1994) (Gernsbacher et al 2005).

Wing and Potter (2002) felt that only about three-quarters of children diagnosed with DSM-IV autism would be diagnosed by the stricter criteria of Leo Kanner from the 1940s.

REASON 5. Changes in society have produced the increase in numbers of children with mental disorders. Children are behaving as they have always behaved, but that it is no longer accepted as normal. These changes include how



children are perceived in modern society as well as moral panic about the state of children.

REASON 6. Official policy changes have produced the increasing number of cases. For example, educational policies giving special assistance to children with a formal diagnosis has led parents to seek such labels to gain that extra assistance.

REASON 7. The increase in cases is a facet of data collection and methodological problems involved. There is no genuine increase in number of cases.

For example, in the USA, under the Individuals With Disabilities Education Act (IDEA) (passed in 1991), schools provide an annual count of the number of children with disabilities. It is this data that are reported as showing dramatic increases. But autism as a category was not introduced until 1991-2, and any new category will show large increases until the use of the category becomes familiar to users. For example, the category of "traumatic brain injury" was introduced at the same time as autism, and this showed a dramatic initial increase (Gernsbacher et al 2005).

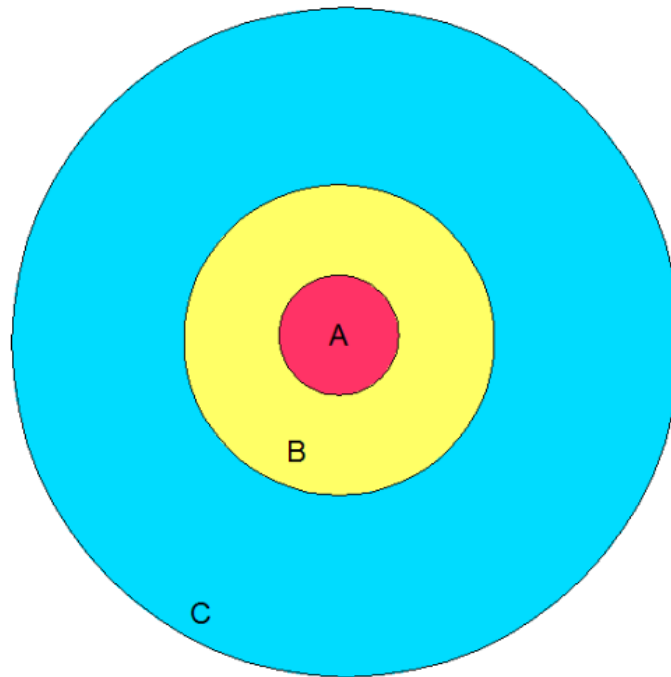
#### **7.4. MAKING SENSE OF THE FIGURES**

Gernsbacher et al (2005) are clear in their conclusion: "no scientific evidence indicates that the increase in the number of diagnosed cases of autism arises from anything other than intentionally broadened diagnostic criteria, coupled with deliberately greater public awareness and conscientiously improved case finding" (p55).

In the situation with any mental disorder, there will be clear cases of individuals who are ill, and, in some respects, it does not need professionals to spot them. They are so obvious by the severity of symptoms and/or distress that ordinary individuals can see them. This is the "hard core" (category A in figure 7.1).

Then there are individuals in category B in figure 7.1 who are not so clear-cut. They show less severe symptoms, and there will be some disagreement among professionals over diagnosis. For me, category C is the interesting one. These are individuals who are "jumping on the bandwagon". What I mean is that the expectations of society encourage individuals (or parents) to be seen as suffering from the condition. For example, greater expectations for children to sit still for longer periods

at school, and when they do not, it is seen as a problem like hyperactivity. This category of individuals have an entirely socially constructed version of the disorder. So the changes in expectations in society will influence the size of category C, in particular, and the consequent number of cases diagnosed.



- A = "Hard core"; no doubt over diagnosis
- B = Some concern and disagreement over diagnosis
- C = "Jumping on the bandwagon"; socially constructed by social norms and expectations of behaviour

Figure 7.1 - Categories of individuals with mental disorders.

In terms of ADHD, two avenues to increased diagnosis (category C) can be seen in Western societies, I would argue. Firstly, among "working class" children, ADHD is a means of social control for their "difficult" behaviour - their failure to conform to social norms at school. But for "middle-class" children, ADHD is a label to justify a child not being "number one". In the USA, in particular, where the emphasis is upon winners and being the best,

how do parents deal with a child who is not top of the class? If "first is everything and second is nothing", there are a lot of unhappy (unsuccessful) people.

## 7.5. REFERENCES

- Atladottir, H.O et al (2007) Time trends in reported diagnosis of childhood neuropsychiatric disorders: A Danish cohort study Archives of Pediatric and Adolescent Medicine 161, 193-198
- APA (1968) Diagnostic and Statistical Manual of Mental Disorders (2nd ed) - DSM-II Washington DC: American Psychiatric Association
- APA (1980) Diagnostic and Statistical Manual of Mental Disorders (3rd ed) - DSM-III Washington DC: American Psychiatric Association
- APA (1987) Diagnostic and Statistical Manual of Mental Disorders (3rd ed revised) - DSM-III-R Washington DC: American Psychiatric Association
- Centers for Disease Control and Prevention (2007a) Prevalence of autism spectrum disorders - Autism and Developmental Disabilities Monitoring Network, six sites, US, 2000 Morbidity and Mortality Weekly Report 56, SS-1, 1-11
- Centers for Disease Control and Prevention (2007b) Prevalence of autism spectrum disorders - Autism and Developmental Disabilities Monitoring Network, fourteen sites, US, 2000 Morbidity and Mortality Weekly Report 56, SS-1, 12-29
- Chakrabati, S & Fombonne, E (2005) Pervasive developmental disorders in preschool children: Confirmation of high prevalence American Journal of Psychiatry 162, 1133-1141
- Clements, S & Peters, J (1962) Minimal brain dysfunction in the school-age child Archives of General Psychiatry 6, 185-197
- Comings, D.E (1996) The Gene Bomb Duarte, CA: Hope Press
- Conrad, P (1975) The discovery of hyperkinesis: Notes on the medicalisation of deviant behaviour Social Problems 23, 12-21
- Fombonne, E (2003) The prevalence of autism Journal of the American Medical Association 289, 1, 87-89
- Gernsbacher, M.A et al (2005) Three reasons not to believe in an autism epidemic Current Directions in Psychological Science 14, 2, 55-58
- Grinker, R.R (2007) Unstrange Minds: Remapping the World of Autism New York: Basic Books
- Kaye, J.A et al (2001) MMR vaccine and the incidence of autism recorded by GPs: A time trend analysis British Medical Journal 24/2, 460-463
- Kutchins, H & Kirk, S (1997) Making Us Crazy New York: Free Press
- Laufer, M et al (1957) Hyperkinetic Impulse Disorder in children's behaviour problems Psychosomatic Medicine 19, 38-49
- Lillienfeld, S.O & Arkowitz, H (2007) Autism: An epidemic? Scientific American Mind April/May, 82-83
- Madsen, K.M (2007) Increasing incidence of some neuropsychiatric disorders in Danish children Evidence-Based Mental Health November, p127
- Mayes, R & Rafalovich, A (2007) Suffer the restless children: The evolution of ADHD and paediatric stimulant use 1900-1980 History of Psychiatry 18, 4, 435-457

Sax, L & Kautz, K.J (2003) Who first suggests the diagnosis of attention deficit/hyperactivity disorder? Annals of Family Medicine 1, 171-174

Schneider, H & Eisenberg, D (2006) Who receives a diagnosis of attention-deficit/hyperactivity disorder in the US elementary school population? Pediatrics 117, 4, e601-609

Schrag, P & Divoky, D (1975) The Myth of the Hyperactive Child and Other Means of Child Control New York: Pantheon

Shattuck, P.T (2006) The contribution of diagnostic substitution to the growing administrative prevalence of autism in US special education Pediatrics 117, 4, 1028-1038

Shorter, E (1997) A History of Psychiatry New York: John Wiley

Still, G (1902) The Goulstonian Lectures on some abnormal physical conditions in children Lancet 26/4, 1163-1167

Stone, M.H (1998) Healing the Mind London: Pimlico

Tredgold, A (1972) Mental Deficiency (Amentia) (4th ed) New York: William Wood & Co

Wing, L & Potter, D (2002) The epidemiology of autistic spectrum disorder: Is the prevalence rising? Mental Retardation and Developmental Disabilities Research Review 8, 151-162

## **8. RELIABILITY AND VALIDITY IN MEASURING MENTAL DISORDERS**

- 8.1. Reliability and validity
- 8.2. Example with pathological gambling
- 8.3. Appendix 8A - reliability and diagnosis
- 8.4. References

### **8.1. RELIABILITY AND VALIDITY**

The criteria for diagnosing mental disorders are hypothetical constructs which have to be operationalised before use. This means that the hypothetical constructs are converted into measurable behaviours. But these measures should show reliability and validity for accuracy purposes.

Reliability shows that the measure is consistent - both within itself (internal reliability) and across time (external reliability) (table 8.1) (appendix 8A). Validity relates to the issue of whether the questionnaire or test measures what it claims to measure (table 8.2) (box 8.1).

- Internal - correlation of scores on individual items in the questionnaire. Cronbach's alpha (Cronbach 1951) is a statistical technique that calculates all the possible correlations between items.
- External - test-retest reliability is the most common type used, and it is the correlation of the same individual's scores on the same test at two different points in time.

Table 8.1 - Types of reliability.

- Face or content validity - on the surface of it, do the questions appear valid.
- Construct validity - the scores on the measuring instrument are correlated with expected behaviours (table 8.3).
- Concurrent validity - the correlation of scores on two questionnaires of the same or similar behaviour.

Table 8.2 - Types of validity.

Face or content validity: "Do you feel sad often?" (high validity); "Do you like chocolate?" (low).

Construct validity (expected behaviours): anti-depressant use, suicidal thoughts (high); type of films watched, time go to bed (low validity).

Size of hand as measure of depression: reliable (ie: consistent measures) but not valid.

Box 8.1 - Examples of valid measures for depression.

BEHAVIOUR	PREDICTION	TEST A	TEST B
High level of fear	Positive correlation	High positive correlation	Low positive correlation
Worry about things that unlikely to happen	Positive correlation	High positive correlation	No correlation
Relaxed before examination	Negative correlation	High negative correlation	Positive correlation
Likes chocolate	No correlation	No correlation	Low positive correlation

Table 8.3 - Example of high (test A) and low (test B) validity tests for anxiety.

## 8.2. EXAMPLE WITH PATHOLOGICAL GAMBLING

Stinchfield (2003) recruited 803 members of the general population in the Minnesota area of the USA and 259 individuals in gambling treatment programmes in that area to aid in studying the DSM-IV (APA 1994) criteria of pathological gambling.

DSM-IV has ten diagnostic criteria which were paraphrased into nineteen self-report yes/no questions (eg: "restless or irritable when attempting to cut down or stop gambling"). Answering "yes" on five or more items was the cut-off point for a diagnosis of pathological gambling.

Cronbach's alpha was calculated as 0.98 for the scores of the combined groups.

Construct validity was established by comparing the mean scores of the two groups. A measure of pathological gambling that does not distinguish (ie: significant difference) between the different groups (gamblers and general population) is not valid. The mean score of the gamblers (8.5) was significantly different to the general population (0.1) ( $p < 0.01$ ; unrelated t-test).

Concurrent validity was established by comparing the

scores on the gambling questionnaire with the South Oaks Gambling Screen (SOGS) (Lesieur and Blume 1987). The separate correlations for each group were significant at  $p < 0.01$  - gamblers ( $r = 0.75$ ) and general population ( $r = 0.77$ ).

The accuracy of the cut-off point (score of five) was tested with the gamblers. How many of them were classed as pathological gamblers using this cut-off point, and how many were missed (table 8.4)? Table 8.5 shows that a cut-off point of four was slightly more accurate than five. The lower cut-off point missed less cases (false negatives), but did have more false positives (wrongly diagnosed). For the general population group, 99.1% of respondents scored below five compared to 5% of gamblers.

GAMBLER:	YES	NO
MEASURING DEVICE SAYS: GAMBLER	Hit	False positive
NOT GAMBLER	False negative	Hit

Table 8.3 - Four possible situations of accuracy.

	CUT-OFF = 5	CUT-OFF = 4
Hit	0.98	0.99
False positive	0.004	0.01
False negative	0.05	0.03

Table 8.4 - Cut-off scores of four and five and accuracy.

### 8.3. APPENDIX 8A - RELIABILITY AND DIAGNOSIS

Reliability in relation to diagnosis can mean something slightly different to it applied to measurement devices. Reliability of a diagnostic category is shown if the independent raters diagnose the same disorder based upon the same symptoms, or if the same diagnosis is given for the same symptoms after a period of time.

For example, Phillips et al (1998), in distinguishing between depressive personality disorder and depression, used separate psychiatrists to diagnose each participant. The first interviewer used the Diagnostic Interview for Depressive Personality (using DSM-IV criteria) (Gunderson et al 1994). The second interviewer used the Structured Clinical Interview for DSM-III-R (SCID) (Spitzer et al 1992), the Diagnostic

Interview for Personality Disorders, Revised (Zanarini et al 1987) (using DSM-III-R criteria; APA 1987), and the Hamilton Depression Rating Scale (HDRS) (Hamilton 1960). The measures were administered one year later.

Based on the Diagnostic Interview for Depressive Personality administered at two different times by two different interviewers, there was 55% agreement in diagnosis at time point A and point B.

#### 8.4. REFERENCES

APA (1987) Diagnostic and Statistical Manual for Mental Disorders (3rd ed revised) DSM-III-R Washington DC: American Psychiatric Association

APA (1994) Diagnostic and Statistical Manual for Mental Disorders (4th ed) DSM-IV Washington DC: American Psychiatric Association

Cronbach, L.J (1951) Coefficient alpha and the internal structure of tests Psychometrika 16, 297-334

Gunderson, J.G et al (1994) The Diagnostic Interview for Depressive Personality American Journal of Psychiatry 151, 1300-1304

Hamilton, M (1960) A rating scale for depression Journal of Neurology, Neurosurgery and Psychiatry 23, 56-62

Lesieur, H.R & Blume, S.B (1987) The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers American Journal of Psychiatry 144, 1184-1188

Phillips, K.A et al (1998) Reliability and validity of depressive personality disorder American Journal of Psychiatry 155, 1044-1048

Spitzer, R.L et al (1992) The Structured Clinical Interview for DSM-III-R (SCID), I: History, rationale, and description Archives of General Psychiatry 49, 624-629

Stinchfield, R (2003) Reliability, validity, and classification accuracy of a measure of DSM-IV diagnostic criteria for pathological gambling American Journal of Psychiatry 160, 180-182

Zanarini, M.C et al (1987) The Diagnostic Interview for Personality Disorders: Inter-rater reliability and test-retest reliability Comparative Psychiatry 28, 467-480



## **9. EATING DISORDER EXAMINATION: AN INVESTIGATOR-BASED INTERVIEW**

The Eating Disorder Examination (EDE) (Cooper and Fairburn 1987) is a commonly used investigator-based interview for eating disorders <sup>14</sup>. This type of interview means that the interviewer scores the answers given based on their judgment, including asking additional questions, rather than just the answers given (table 9.1).

INVESTIGATOR-BASED INTERVIEW	INVESTIGATOR-SCORED INTERVIEW
<ul style="list-style-type: none"><li>• Interviewer scores questionnaire based on answers given and own judgment.</li><li>• Tends to use semi-structure interviewing which allows flexibility depends upon interviewee.</li><li>• Interviewer able to use expertise when scoring answers.</li><li>• Able to include judgments beyond what person says, particularly if they lack insight or are inconsistent.</li></ul>	<ul style="list-style-type: none"><li>• Interviewer scores questionnaire based on answers given only.</li><li>• Tends to use structured interviewing where the same questions are asked in exactly same way with each person.</li><li>• Good for standardisation and thus comparability of data, particularly with different interviewers.</li><li>• Does not include subjective judgments of interviewer, only what interviewee says.</li></ul>

Table 9.1 - Investigator-based and investigator-scored interviews.

"The interviewer and participant together should be trying to obtain an accurate picture of the participant's current eating behaviour and attitudes" (Fairburn et al 2008).

The latest version, EDE 16.0D (Fairburn et al 2008) (box 9.1), lasts between 45-90 minutes, and focuses mainly on the last 28 days. To aid accurate recall, details of the participant's life in the last month is used to structure events (eg: returned from holiday four weeks ago).

---

<sup>14</sup> There is a version specifically designed for children, and a self-report version.

Over the past four weeks:

- Have you "felt fat"?
- Have you felt uncomfortable about others seeing your body, for example, in communal changing rooms, when swimming, or when wearing clothes that show your shape? What about your partner or friends seeing your body?
- Have you been afraid that you might gain weight?
- Have you been dissatisfied with your overall shape (your figure)? What has this been like?
- Have you taken laxatives as a means of controlling your weight or shape?
- Have you wanted your stomach to be empty? has this been to influence your shape or weight, or to avoid triggering an episode of overeating?

(Source: Fairburn et al 2008)

#### Box 9.1 - Example of questions on EDE 16.0D.

The data collected takes two forms:

i) Frequency - how often of certain behaviours (table 9.2).

- 0 - Absence of the feature
- 1 - Feature present on 1 to 5 days
- 2 - Feature present of 6 to 12 days
- 3 - Feature present on 13 to 15 days
- 4 - Feature present on 13 to 15 days
- 5 - Feature present almost every day (23 to 27 days)
- 6 - Feature present every day
- 8 - If, despite adequate questioning, it is impossible to decide upon rating.
- 9 - Missing values or not applicable

(Source: Fairburn et al 2008)

#### Table 9.2 - Frequency ratings used in EDE 16.0D.

ii) Severity - how much of certain behaviours (table 9.3).

- 0 - Absence of the feature
- 1 - Feature almost, but not quite absent
- 2
- 3 - Severity midway between 0 and 6
- 4
- 5 - Severity almost meriting a rating of 6
- 6 - Feature present to an extreme degree
  
- 8 - If, despite adequate questioning, it is impossible to decide upon rating.
- 9 - Missing values or not applicable

(Source: Fairburn et al 2008)

Table 9.3 - Severity ratings used in EDE 16.0D.

Participants are given an overall (global) score, and four sub-scale scores. The sub-scales focus on specific characteristics of eating disorders:

- Restraint (over eating);
- Eating concern (eg: fear of losing control over eating);
- Shape concern;
- Weight concern.

## REFERENCES

Cooper, Z & Fairburn, C.G (1987) The Eating Disorder Examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders International Journal of Eating Disorders 6, 1-8

Fairburn, C.G et al (2008) Eating Disorder Examination (edition 16.0D). In Fairburn, C.G (ed) Cognitive Behaviour Therapy and Eating Disorders New York: Guilford Press

## **10. NEW MENTAL DISORDERS: PROLONGED GRIEF DISORDER**

There are a number of disorders proposed in the appendices of DSM-IV with the aim of inclusion in the future DSM-V.

One such example is prolonged grief disorder (PGD). Grief after bereavement is normal, but this disorder attempts to highlight abnormal grief. The distinct characteristic of PGD is yearning ("physical and emotional suffering as a result of the desired, but unfulfilled, reunion with the deceased"; Prigerson et al 2009).

Diagnosis also requires the presence of at least five from nine other symptoms which are experienced at least daily or experienced to a disabling degree. The symptoms must be present at high levels at least six months after the death (Prigerson et al 2009):

- Feeling emotional numbness or absence of emotion since the loss;
- Feeling stunned, dazed or shocked by the loss;
- Feeling life is meaningless, unfulfilled or empty since the loss;
- Experiencing mistrust of others;
- Bitterness and anger related to the loss;
- Difficulty accepting the loss;
- Confusion about role in life or diminished sense of self (ie: feeling that part of self has died);
- Avoidance of reminders of the reality of the loss;
- Difficulty moving on with life (eg: making new friends).

A number of risk factors for PGD have been proposed (Prigerson et al 2009):

- History of childhood separation anxiety;
- Controlling parents;
- Parental abuse or death;
- Close relationship to deceased;
- Insecure attachment style;
- Marital dependency;
- Lack of preparation for the death.

There is concern that PGD overlaps with other disorders like major depression, but distinctive features have been found. For example, yearning does not appear important in bereaved individuals with depression and anxiety, whereas sadness and anxiety are (Prigerson et al 2009).

Prigerson et al (2009) collected data from the Yale Bereavement Study (YBS) which was a longitudinal study of community-dwelling bereaved older adults in Connecticut, USA. The 317 participants were interviewed at approximately six months (baseline), 11 months, and twenty months after the loss using the Inventory of Complicated Grief - Revised (ICG-R). During a structured interview, the interviewer rated each PGD symptom on a scale of 1-5 (with a score of 4 or 5 viewed as a problem). Other questionnaires measured suicidal thoughts and behaviours ("Yale Evaluation of Suicidality"), everyday functioning ("Established Populations for Epidemiological Studies of the Elderly"), and quality of life ("Medical Outcomes Short-Form").

At 6-12 months post-loss, 3.3% of individuals were diagnosed with PGD. These individuals were eight times more likely to have major depression, anxiety or Post-Traumatic Stress Disorder, five times more likely to have suicidal thoughts, and poor quality of life, and twice as likely to struggle to function with everyday life at 12-24 months post-loss (table 10.1).

	PGD DIAGNOSIS	NO PGD DIAGNOSIS
Depression, anxiety, PTSD	28.6	3.4
Suicidal thoughts	57.1	10.1
Functional disability	71.4	35.9
Poor quality of life	83.3	14.7

(After Prigerson et al 2009)

Table 10.1 - Percentage of individuals showing certain behaviours at 12-24 months post-loss based on diagnosis of PGD.

Implicit in the criteria for PGD are social norms and expectations, most prominently that individuals should have recovered "normal" functioning after six to twelve months. What happens if individuals are so grief-stricken that it takes years to recover from the loss. That is defined as abnormal or pathological. PGD is a categorical means of defining normality. Yet individuals are different and loss will have consequences in different ways for different lengths of time.

## REFERENCE

Prigerson, H et al (2009) Prolonged grief disorder: Psychometric validation of criteria proposed for DSM-V and ICD-11 PLoS Medicine 6, 8, e1000121  
(<http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.1000121>)

## **11. ADULT PSYCHIATRIC MORBIDITY SURVEY**

### **2007: METHODOLOGY**

The "Adult Psychiatric Morbidity Survey" (APMS) (McManus et al 2009) was based on data collected in 2007 (thus APMS 2007) by the NatCen (National Centre for Social Research). It continued the series of surveys in 1993 and 2000 by the Office for National Statistics (ONS). Though a non-governmental organisation collected the data, the APMS 2007 is classed as official statistics.

It was a cross-sectional study <sup>(1)</sup> designed to be representative of the population living in private households in England <sup>(2)</sup>. The sampling method was "multi-stage stratified probability sampling". This means that the sample was designed to mirror the general population (stratified) using a number of stages, and then specific individuals were chosen for interview (probabilistic) <sup>(3)</sup>.

Stage 1 of sampling was the selection of primary sampling units (PSUs). Using Royal Mail postcode and delivery information, postal sectors of about 2500 households were defined. The postcode sectors were stratified based on population census data about socio-economic status and proportion of households without a car. From this process, 519 postal sectors were selected.

Stage 2 involving the sampling of households within these postal sectors. Twenty-eight households were randomly sampled from each sector (total = 14 532 households <sup>(4)</sup>), and one adult sixteen years and over was randomly chosen in each household <sup>(5)</sup>.

Of the individuals chosen, 57% (7461) agreed to participate and fully completed the interview <sup>(6)</sup> (figure 11.1).

An advanced letter was sent to each household with details of the survey, which was titled the "National Study of Health and Wellbeing" <sup>(7)</sup>. The interview was set to take 1.5 hours to complete (though it could take three hours) (Scholes et al 2009). Some of the questions were self-report using the interviewer's laptop <sup>(8)</sup>. The respondents were given a small value gift voucher (£5-10) for completion of the interview <sup>(9)</sup>. They were left with details of a helpline and a "thank you letter" arrived soon afterwards.

The interviewers were given one day's training on using the survey, and on responding to participant's distress <sup>(10)</sup>. Written instructions were also provided as the interviews took place over one year <sup>(11)</sup>.

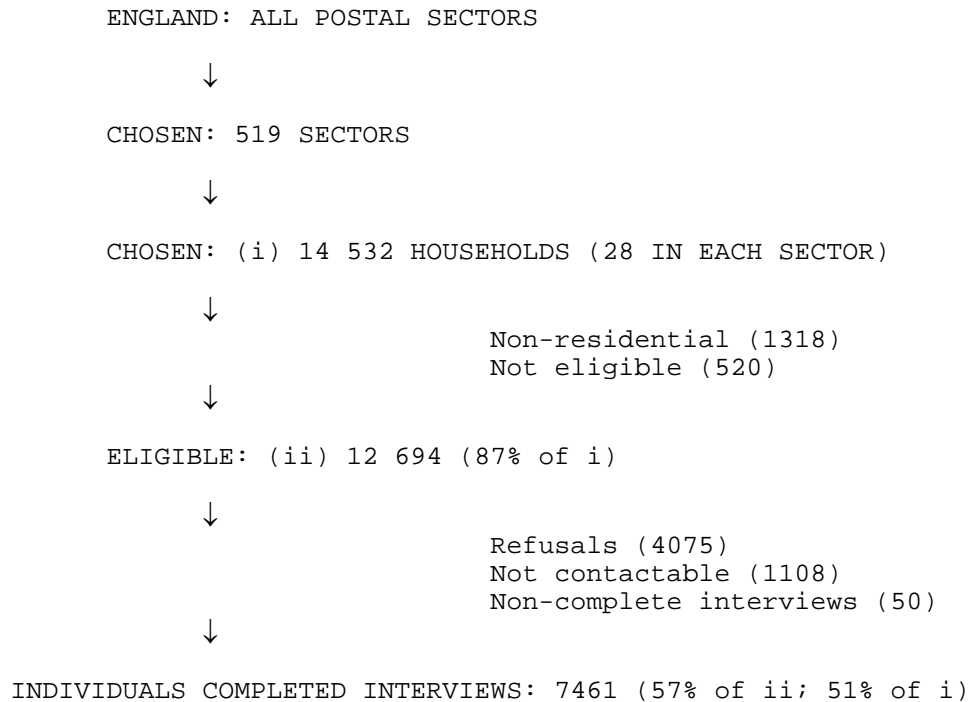


Figure 11.1 - Sampling stages and number involved.

A second set of more detailed interviews was conducted with 630 individuals <sup>(12)</sup>, who responded positively to questions on psychosis, Asperger syndrome or personality disorders.

Table 11.1 gives an example of the results from APMS 2007.

## EVALUATIVE COMMENTS

(1) A cross-sectional study compares different groups at one point in time, like age, gender or ethnicity groups. It shows the difference between them, but it cannot explain behaviour over time as with a longitudinal study. A cross-sectional study is a one-shot ("snap-shot") in time survey.

(2) Individuals living in institutions (2% of 16 year-olds and over in England) and outside private households (eg: homeless) were not included. Such groups are known to have poorer mental health than the general population (Scholes et al 2009).

(3) Stratified sampling produces a better representation of the general population than random sampling, but it is time consuming and complex to design.

	ALL	MEN	WOMEN
ALL	16.2	12.5	19.7
Age groups:			
16-24	17.5	13.0	22.2
25-34	18.6	14.6	23.0
35-44	17.3	15.0	19.5
45-54	19.9	14.5	25.2
55-64	14.1	10.6	17.6
65-74	10.6	7.5	13.4
75+	9.9	6.3	12.2
Ethnicity:			
White	/	11.9	19.2
Black		16.3	25.3
South Asian		11.3	23.4
Other		19.4	21.1
Marital status:			
Married	/	10.1	16.3
Cohabiting		14.0	21.6
Single		14.8	24.6
Widowed		10.4	17.4
Divorced		27.7	26.6
Separated		10.5	33.0
Equivalised household income:			
Highest	/	8.8	18.1
2nd		8.6	13.1
3rd		10.1	20.1
4th		16.2	24.0
Lowest		23.5	25.1

(Source: Deverill and King 2009)

(/ = not given in Deverill and King 2009)

(Common mental disorders include mixed anxiety and depressive disorder, generalised anxiety disorder, depressive episode, all phobias, obsessive-compulsive disorder, and panic disorder)

Table 11.1 - Prevalence (%) of any common mental disorder among adults (16 years and over) in England in past week in 2007.

(4) 87% of addresses were eligible (12 694). The others were ineligible because, for example, not residential addresses.

(5) With the random sampling elements, every household had an equal chance of being chosen, and within the households, each adult member had an equal chance of being interviewed.

(6) Respondents and non-respondents to a survey may not be the same type of people, and thus a bias sample could occur (eg: more individuals with mental illness are non-respondents). Government statisticians weight the data during analysis to take account of non-respondents, and to make the results representative of the general



population. Statistical techniques like logistic regression modelling are used.

(7) It was felt that this title was better than "psychiatric morbidity" or something involving mental illness. There is a small element of deception beforehand, though the participants may realise the true purpose when questioned.

(8) The aim was to interview the respondent alone, but if the selected person was not capable of such interviews, a shortened "proxy interview" (n = 58) was conducted with a close relative or carer.

(9) There is a debate about whether respondents should receive remuneration for their participation:

- Remuneration could change the nature of the interviewer-interviewee relationship, and consequently the answers given;
- The remuneration here was a token amount given afterwards. The respondents were not expecting it, and had agreed to participate as volunteers;
- The token amount may have felt like an insult after a long interview. But how much is an appropriate payment, then?

(10) With a structured interview, as used here, it is important to have standardised interviews which is the reason for the training. Also supervisors attended the early interviews and sampled 10% of interviews overall by telephoning interviewees.

(11) Surveys can be used to collect vast amounts of data, but they are time consuming. There may be differences in society between the first interviews and the last ones a year later as well as the interviewees changing. For example, interviews conducted before something like 11 September 2001 ("9/11") may produce different results to those conducted afterwards.

(12) Of 7461 main interviewees, 4050 agreed to a second interview. From this group a sample of 849 was made and 630 completed (74% response rate).

## REFERENCES

Deverill, C & King, M (2009) Common mental disorders. In McManus et al op cit

McManus, S et al (2009) (eds) Adult Psychiatric Morbidity, 2009: Results of a Household Survey London: National Centre for Social Research

Scholes, S et al (2009) Methods. In McManus et al op cit